

BAB I

PENDAHULUAN

1.1. Latar Belakang

Plagiarisme dalam penelitian dapat saja terjadi karena ketidaksengajaan ataupun disengaja. Plagiarisme tentunya merupakan tindakan yang melanggar hak cipta dan termasuk perbuatan yang merugikan orang lain. Dalam pengajuan judul penelitian, contohnya untuk penelitian tugas akhir tidak sedikit mahasiswa yang berulang kali mengajukan judul dikarenakan judul yang diajukan telah ada sebelumnya, sehingga di tolak karena dianggap melakukan plagiat.

Menentukan diterima atau tidaknya sebuah judul penelitian seperti penelitian tugas akhir yang sekarang ini dilakukan adalah dengan melalui kontrol dan seleksi dari para dosen atau pembimbing dalam mengetahui secara dini apakah judul penelitian tersebut sudah ada atau belum. Terkadang kemampuan dosen di dalam melakukan kontrol dan penyeleksian masih terkendala dengan harus mengecek dan mengetahui dengan kemampuan daya ingat dari masing-masing dosen atau pembimbing yang mungkin terbatas, sehingga terkadang satu dua judul lolos dan luput dari pantauan sehingga terjadi duplikasi judul.

Duplikasi judul merupakan bentuk plagiarisme yang umum terjadi dalam penulisan tugas akhir (Tri Nugroho I, 2017). Untuk mengatasi masalah tersebut, maka diperlukan suatu sistem yang bisa mencari seberapa besar persentase kemiripan judul penelitian yang diajukan mahasiswa dengan judul penelitian yang telah ada sebelumnya. Data judul penelitian yang telah ada di *google scholar* yang mencakup jurnal-jurnal *online* dari publikasi ilmiah dapat dimanfaatkan

untuk membantu dalam mendapatkan judul lain yang sudah ada sebagai referensi atau pembanding.

Penerapan *web scraping* dengan CURL (*Client URLs*) dan *simple HTML DOM parser* diharapkan dapat membantu untuk pengambilan data judul pembanding dari data judul penelitian yang telah ada di *google scholar* karena sampai saat ini tidak tersedia API untuk mendapatkan data judul penelitian dari *google scholar*. CURL merupakan sebuah *library* dan baris perintah untuk melakukan transfer data ke dan dari *server*. CURL digunakan untuk metode pengambilan data dari situs (Imran Khalid, 2005). *Simple HTML DOM parser* membantu memanipulasi elemen HTML. Tidak hanya terbatas pada class HTML yang valid, tetapi juga dapat bekerja dengan kode HTML yang tidak termasuk validasi W3C. Pada HTML DOM dapat ditemukan unsur menurut id, kelas, tag, dan banyak lagi. Elemen DOM juga dapat ditambahkan, dihapus atau diubah (Janjic, 2011).

Algoritma *winnowing* dapat digunakan untuk mencari persentase kemiripan teks judul penelitian yang diajukan dengan judul penelitian yang telah didapat dari *google scholar*. Algoritma *winnowing* telah memenuhi prasyarat dari algoritma deteksi kemiripan teks yaitu *whitespace insensitivity*, dan algoritma *winnowing* telah memenuhi prasyarat tersebut yaitu membuang seluruh karakter-karakter yang tidak relevan seperti: tanda baca, spasi dan juga karakter lain, sehingga nantinya hanya karakter-karakter yang berupa huruf atau angka yang akan diproses lebih lanjut (Schleimer, Wilkerson, & Aiken, 2003). Algoritma

winnowing menghasilkan tingkat persentase yang lebih baik dan waktu proses yang lebih cepat dibandingkan algoritma *rabin-karp* (Alamsyah, 2017).

Berdasarkan permasalahan tersebut, untuk mengurangi plagiarisme dan mendeteksi dini pengajuan judul penelitian mahasiswa maka akan dilakukan penelitian yang berjudul “**Implementasi *Web Scraping* Pada *Google Scholar* Untuk Deteksi Dini Pengajuan Judul Penelitian Menggunakan Algoritma *Winnowing*”**”.

1.2. Rumusan Masalah

Dari latar belakang diatas, maka masalah yang dihadapi adalah :

1. Bagaimana menerapkan *web scraping* dengan *CURL* dan *simple HTML DOM parser* yang digunakan untuk melakukan pengambilan data dari *google scholar* pada aplikasi deteksi dini pengajuan judul penelitian mahasiswa?
2. Bagaimana menerapkan algoritma *winnowing* untuk mencari persentase kemiripan judul penelitian yang diajukan dengan judul penelitian yang telah ada di *google scholar* pada aplikasi deteksi dini pengajuan judul penelitian mahasiswa

1.3. Batasan Masalah

Beberapa hal yang menjadi batasan masalah adalah sebagai berikut :

1. Aplikasi deteksi dini pengajuan judul penelitian mahasiswa ini berbasis web.

2. Aplikasi hanya mencari persentase kemiripan judul penelitian yang diajukan mahasiswa dengan data judul yang telah ada di *google scholar*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Mengimplementasikan *web scraping* dengan *CURL* dan *simple HTML DOM parser* yang digunakan untuk melakukan pengambilan data dari *google scholar* pada aplikasi deteksi dini pengajuan judul penelitian mahasiswa.
2. Menerapkan algoritma *winnowing* untuk mencari persentase kemiripan judul penelitian yang diajukan dengan judul penelitian yang telah ada di *google scholar* pada aplikasi deteksi dini pengajuan judul penelitian mahasiswa.

1.5. Manfaat Penelitian

Membantu mengurangi terjadinya duplikasi judul penelitian dan plagiarisme dengan mengetahui persentase kemiripan judul penelitian yang diajukan mahasiswa atau peneliti dengan judul yang sudah ada di *google scholar*.

1.6. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah :

1. Studi Literatur : Studi literatur dilakukan dengan mengumpulkan informasi yang dijadikan sebagai acuan untuk penelitian ini, yang berkaitan dengan *web scraping*, *CURL*, *simple HTML DOM parser*,

google scholar, dan algoritma *winning*. Baik dari buku, jurnal dan melalui penemuan literatur internet studi kepustakaan.

2. Identifikasi Masalah : Mempelajari dan memahami masalah yang akan diteliti, sehingga pada proses penelitian tidak menyimpang atau keluar dari batasan masalah.
3. Pengembangan perangkat lunak : Metode pengembangan perangkat lunak yang digunakan adalah metode *Extreme Programming (XP)*. *Extreme Programming (XP)* adalah metodologi pengembangan perangkat lunak yang ditujukan untuk meningkatkan kualitas perangkat lunak dan tanggap terhadap perubahan kebutuhan pelanggan. *Extreme Programming (XP)* memiliki 4 tahapan, yaitu *Planning, Design, Coding* dan *Testing* (Pressman, 2009).

1.7. Sistematika Penulisan

Sistematika penulisan ini akan menjelaskan secara garis besar pada setiap bab sehingga didapatkan gambaran isi pada setiap bab yang ditulis.

BAB I PENDAHULUAN

Bab ini berisikan tentang gambaran secara garis besar tentang isi laporan, yang didalamnya memuat Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Metodologi dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini membahas tentang teori-teori yang berhubungan dengan penelitian dan berisi jurnal-jurnal yang relevan sehingga menjadi acuan dilakukan penelitian ini.

BAB III METODOLOGI

Bab ini menjelaskan tentang metode yang digunakan dalam pengembangan sistem perangkat lunak dengan memuat konsep dari metode yang digunakan, analisis kebutuhan, masukan maupun keluaran, perancangan, serta uraian-uraian yang berkaitan dengan laporan yang akan dibahas sesuai yang dibutuhkan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil proses dari analisis serta perancangan yang telah dilakukan dan bagaimana hasil tersebut dicapai dari setiap aktivitas yang dilakukan selama penelitian seperti implementasi dan pengujian perangkat lunak, pengujian yang dilakukan untuk mengetahui kekurangan maupun kelebihan dari aplikasi yang dibutuhkan.

BAB V KESIMPULAN DAN SARAN

Bab ini memuat simpulan yang merupakan rangkuman dari hasil penelitian, dan memuat saran yang berisi saran-saran yang perlu diperhatikan berdasarkan keterbatasan yang ditemukan dalam pembuatan perangkat lunak tersebut.